(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 779 198**

(51) Int. CI.:

H04R 1/40 (2006.01)
H04R 3/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **02.12.2011 PCT/EP2011/071600**

(87) Fecha y número de publicación internacional: **07.06.2012 WO12072787**

(96) Fecha de presentación y número de la solicitud europea: **02.12.2011 E 11808175 (1)**

(97) Fecha y número de publicación de la concesión europea: **08.01.2020 EP 2647221**

(54) Título: **Aparato y procedimiento para la adquisición espacialmente selectiva del sonido mediante triangulación acústica**

(30) Prioridad:

03.12.2010 US 419720 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
14.08.2020

(73) Titular/es:

FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (100.0%)
Hansastrasse 27c
80686 München, DE

(72) Inventor/es:

HERRE, JÜRGEN;
KÜCH, FABIAN;
KALLINGER, MARKUS;
DEL GALDO, GIOVANNI y
GRILL, BERNHARD

(74) Agente/Representante:

SALVÀ FERRER, Joan

DESCRIPCIÓN

Aparato y procedimiento para la adquisición espacialmente selectiva del sonido mediante triangulación acústica

[0001] La invención se refiere al procesamiento de audio y, en particular, a un aparato para capturar información de audio desde una ubicación objetivo. Además, la solicitud se relaciona con la adquisición de sonido espacialmente selectiva mediante triangulación acústica.

[0002] La adquisición de audio espacial tiene como fin capturar un campo sonoro completo que está presente en una sala de grabación o solo ciertos componentes deseados del campo sonoro que son de interés para la aplicación en uso. Por ejemplo, en una situación en la que varias personas tienen una conversación en una sala, puede ser de interés capturar el campo sonoro completo (incluyendo sus características espaciales) o solo una señal que produce un interlocutor dado. Esto último permite aislar el sonido y aplicar al mismo un procesamiento específico tal como amplificación, filtrado, etc.

[0003] Hay un número de procedimientos conocidos para capturar de manera espacialmente selectiva ciertos componentes de sonido. Estos procedimientos con frecuencia emplean micrófonos con alta direccionalidad o una matriz de micrófonos. La mayoría de los procedimientos tienen la característica común de que el micrófono o la serie de micrófonos está dispuesta con una geometría conocida fija. El espaciamiento entre los micrófonos es lo más pequeño posible para las técnicas de micrófonos coincidentes, mientras que normalmente es de unos pocos centímetros en los otros procedimientos. En lo sucesivo, nos referimos a cualquier aparato para la adquisición direccionalmente selectiva del sonido espacial (por ejemplo, micrófonos direccionales, matrices de micrófonos, etc.) como un formador de haces.

[0004] Tradicionalmente, la selectividad direccional (espacial) en la captura de sonido, es decir, una adquisición de sonido espacialmente selectiva, se puede obtener de varias maneras:

Una manera posible consiste en emplear micrófonos direccionales (por ejemplo, micrófonos cardioide, súper cardioide o de cañón). En ese caso, todos los micrófonos capturan el sonido de manera diferente dependiendo de la dirección de llegada (DOA) con respecto al micrófono. En algunos micrófonos, este efecto es menor, ya que capturan el sonido casi independientemente de la dirección. Estos micrófonos se denominan micrófonos omnidireccionales. Por lo general, en esos micrófonos, hay un diafragma circular ajustado a un pequeño encierro hermético; véase, por ejemplo,

[Ea01] Eargle J. "The Microphone Book" Focal press 2001.

[0005] Si el diafragma no está sujeto a la cabina y el sonido lo alcanza por igual desde cada lado, su patrón direccional tiene dos lóbulos de igual magnitud. Captura el sonido con igual nivel tanto desde la parte delantera como desde la parte trasera del diafragma, aunque con polaridades inversas. Este micrófono no captura el sonido procedente de direcciones paralelas al plano del diafragma. Este patrón direccional se denomina dipolo o figura ocho. Si la cabina del micrófono omnidireccional no es estanca al aire, sino que se efectúa una construcción especial que permite que las ondas sonoras se propaguen a través de la cabina y alcancen el diafragma, el patrón direccional está en algún punto entre omnidireccional y dipolo (véase [Ea01]). Los patrones pueden constar de dos lóbulos; sin embargo, los lóbulos pueden tener diferentes magnitudes. Los patrones pueden tener también un solo lóbulo; el ejemplo más importante es el patrón cardioide, donde la función direccional D se puede expresar en términos de $D = 0{,}5(1 + \cos(\theta))$, donde θ es la dirección de llegada del sonido (véase [Ea01]). Esta función cuantifica la magnitud relativa del nivel de sonido capturado de una onda plana en un ángulo θ con respecto al ángulo con la mayor sensibilidad. Los micrófonos omnidireccionales se denominan micrófonos de orden cero y otros patrones mencionados anteriormente tales como los patrones dipolares y cardioides, son conocidos como patrones de primer orden. Estos tipos de micrófonos no permiten el modelado arbitrario del patrón, puesto que su patrón de direccionamiento está determinado, casi por completo, por su construcción mecánica.

[0006] También existen algunas estructuras acústicas especiales que se pueden utilizar para generar patrones direccionales más estrechos a los micrófonos que los de primer orden. Por ejemplo, si se conecta un tubo que tiene orificios a un micrófono omnidireccional, se puede crear un micrófono con un patrón direccional muy estrecho. Tales micrófonos se denominan micrófonos de cañón o rifle (véase [Ea01]). Por lo general no tienen respuestas de frecuencia plana y su direccionalidad no se puede controlar después de la grabación.

[0007] Otro procedimiento para la construcción de un micrófono con características direccionales consiste en grabar el sonido con una matriz de micrófonos omnidireccionales o direccionales y aplicar posteriormente el procesamiento de la señal; véase, por ejemplo,
[BW01] M. Brandstein, D. Ward: “Microphone Arrays – Signal Processing Techniques and Applications”, Springer Berlin, 2001, ISBN: 978–3–540–41953–2.

[0008] Existe una variedad de procedimientos para ello. En la forma más sencilla, cuando se graba el sonido

con dos micrófonos omnidireccionales cercanos entre sí y aislados entre sí, se forma una señal de micrófono virtual con una característica dipolar. Véase, por ejemplo,
[Elk00] G. W. Elko: “Superdirectional microphone arrays” in S. G. Gay, J. Benesty (eds.): “Acoustic Signal Processing for Telecommunication”, Capítulo 10, Kluwer Academic Press, 2000, ISBN: 978–0792378143.

[0009] Las señales de micrófonos también pueden ser retardadas o filtradas antes de sumarse entre sí. En la formación de haces, se forma una señal correspondiente a un haz estrecho filtrando cada señal de micrófono con un filtro especialmente diseñado y, a continuación, sumándolas entre sí. Esta “formación de haces por filtrado y suma” está explicada en
[BS01]: J. Bitzer, K. U. Simmer: “Superdirective microphone arrays” en M. Brandstein, D. Ward (eds.): “Microphone Arrays – Signal Processing Techniques and Applications”, Capítulo 2, Springer Berlin, 2001, ISBN: 978–3–540–41953–2.

[0010] Estas técnicas son ciegas a la señal en sí, por ejemplo, no toman en cuenta la dirección de llegada del sonido. Por el contrario, la estimación de la “dirección de llegada” (DOA) es una tarea en sí misma; véase, por ejemplo,
[CBH06] J. Chen, J. Benesty, Y. Huang: “Time Delay Estimation in Room Acoustic Environments: An Overview”, EURASIP Journal on Applied Signal Processing, ID Artículo 26503, Volumen 2006 (2006).

[0011] En principio, se pueden formar muchas características direccionales con estas técnicas. Sin embargo, para formar patrones de sensibilidad arbitrarios espacialmente muy selectivos, se necesita un gran número de micrófonos. En general, todas estas técnicas se basan en las distancias de los micrófonos adyacentes, que son pequeñas en comparación con la longitud de onda de interés.

[0012] Otra manera de obtener la selectividad direccional en la captura de sonido es el filtrado espacial paramétrico. Los diseños normales de formadores de haces que se pueden basar, por ejemplo, en un número limitado de micrófonos y que poseen filtros que no varían en el tiempo en su estructura de filtro y suma (véase [BS01]) habitualmente exhiben selectividad espacial solo limitada. Para aumentar la selectividad espacial, últimamente se han propuesto técnicas de filtrado espacial paramétrico que aplican funciones de ganancia espectral (que varían en el tiempo) al espectro de la señal de entrada. Las funciones de ganancia se diseñan basándose en parámetros que se relacionan con la percepción humana del sonido espacial. Se presenta una estrategia de filtrado espacial en
[DiFi2009] M. Kallinger, G. Del Galdo, F. Küch, D. Mahne y R. Schultz–Amling, “Spatial Filtering using Directional Audio Coding Parameters,” in Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Abr. 2009,
y se implementa en el dominio de parámetros de la Codificación de Audio Direccional (DirAC), una técnica eficiente de codificación espacial. La Codificación de Audio Direccional se describe en
[Pul06] Pulkki, V., "Directional audio coding in spatial sound reproduction and stereo upmixing," en Acta de la 28a Conferencia Internacional de AES, págs. 251–258, Piteå, Suecia, 30 de junio –2 de julio de 2006.

[0013] En la DirAC, se analiza el campo sonoro en una ubicación en la cual se mide el vector de intensidad activo, así como la presión sonora. Estas propiedades físicas se utilizan para extraer los tres parámetros de DirAC: presión sonora, dirección de llegada (DOA) y difusividad del sonido. La DirAC hace uso de la presunción de que el aparato auditivo humano solo puede procesar una dirección por vez y por mosaico de frecuencia. Esta presunción también es aprovechada por otras técnicas de codificación de audio espacial como MPEG Envolvente; véase, por ejemplo:
[Vil06] L. Villemoes, J. Herre, J. Breebaart, G. Hotho, S. Disch, H. Purnhagen y K. Kjörling, “MPEG Surround: The Forthcoming ISO Standard for Spatial Audio Coding,” en 28a Conferencia Internacional de AES, Pitea, Suecia, Junio de 2006.

[0014] La estrategia de filtrado espacial, descrita en [DiFi2009], da lugar a una elección casi libre de la selectividad espacial.

[0015] Una técnica adicional hace uso de parámetros espaciales comparables. Esta técnica se explica en
[Fal08] C. Faller: “Obtaining a Highly Directive Center Channel from Coincident Stereo Microphone Signals”, Acta de la 124a convención de AES, Ámsterdam, Países Bajos, 2008, Preimpresión 7380.

[0016] A diferencia de la técnica descrita en [DiFi2009], en la cual se aplica una función de ganancia espectral a una señal de micrófono omnidireccional, la estrategia planteada en [Fal08] hace uso de dos micrófonos cardioides.

[0017] Las dos técnicas de filtrado espacial paramétrico se basan en las distancias entre los micrófonos, que son pequeñas en comparación con la longitud de onda de interés. Idealmente, las técnicas descritas en [DiFi2009] y [Fal08] se basan en micrófonos direccionales coincidentes.

[0018] Otra manera de obtener la selectividad direccional en la captura de sonido es un filtrado de las señales de micrófonos basadas en la coherencia entre señales de micrófono. En
[SBM01] K. U. Simmer, J. Bitzer y C. Marro: “Post–Filtering Techniques” in M. Brandstein, D. Ward (eds.): “Microphone Arrays – Signal Processing Techniques and Applications”, Capítulo 3, Springer Berlin, 2001, ISBN: 978–3–540–41953–

2,
se describe una familia de sistemas, que emplea al menos dos micrófonos (no necesariamente direccionales) y el procesamiento de su señal de salida se basa en la coherencia de las señales. La presunción subyacente es que el ruido de fondo difuso aparece como partes incoherentes en las dos señales de micrófonos, mientras que una señal de origen aparecerá de manera coherente en estas señales. Basándose en esta premisa, se extrae la parte coherente como señal de origen. Las técnicas mencionadas en [SBM01] fueron desarrolladas debido a que los formadores de haces de filtro y suma con un número limitado de micrófonos tienen escasa capacidad para reducir las señales de ruido difuso. No se efectúan presunciones sobre la ubicación de los micrófonos, ni siquiera es necesario conocer el espaciamiento entre los micrófonos.

[0019] Una limitación importante de las estrategias tradicionales para la adquisición de sonido espacialmente selectiva es que el sonido grabado siempre se relaciona con la ubicación del formador de haces. Sin embargo, en muchas aplicaciones no es posible (o factible) colocar un formador de haces en la posición conveniente, por ejemplo, en un ángulo ventajoso con respecto a la fuente de sonido de interés.

[0020] Por ejemplo, los formadores de haces tradicionales pueden emplear matrices de micrófonos y pueden formar un patrón ("haz") para captar el sonido desde una dirección – y rechazar el sonido de otras direcciones. En consecuencia, no existe posibilidad de restringir la región de la captura de sonido con respecto a su distancia de la matriz de micrófonos de captura.

[0021] Sería sumamente ventajoso contar con un dispositivo de captura que pueda capturar selectivamente el sonido que se origina no solo de una dirección, sino que se limita directamente al origen de una ubicación (punto), de manera similar a la eficiencia con que funcionaría un micrófono de punto próximo en el lugar indicado. Los documentos WO 2007/025033 A2 y WO 2006/006935 A1 describen ejemplos de tal dispositivo y el procedimiento correspondiente.

[0022] El objeto de la presente invención consiste en proporcionar conceptos mejorados para captar información de audio desde una ubicación objetivo. El objeto de la presente invención se alcanza mediante un aparato para capturar información de audio según la reivindicación 1, un procedimiento para computar el sonido según la reivindicación 8 y un programa informático según la reivindicación 9.

[0023] Se proporciona un aparato para capturar información de audio desde una ubicación objetivo. El aparato comprende un primer formador de haces que está dispuesto en un entorno de grabación y que tiene una primera característica de grabación, un segundo formador de haces que está dispuesto en el entorno de grabación y que presenta una segunda característica de grabación y un generador de señales. El primer formador de haces está configurado para registrar la señal de audio de un primer formador de haces y el segundo formador de haces está configurado para grabar una señal de audio del segundo formador de haces cuando el primer formador de haces y el segundo formador de haces se dirigen hacia la ubicación objetivo con respecto a la primera y segunda características de grabación. El primer formador de haces y el segundo formador de haces están dispuestos de tal manera que una primera línea recta virtual, que está definida pasando a través del primer formador de haces y la ubicación objetivo, y una segunda línea recta virtual, que por definición pasa a través del segundo formador de haces y la ubicación objetivo, no son paralelas entre sí. El generador de señales está configurado para generar una señal de salida de audio sobre la base de la señal de audio del primer formador de haces y de la señal de audio del segundo formador de haces de manera que la señal de salida de audio refleje una información de audio de la ubicación objetivo relativamente mayor que la información de audio de la ubicación objetivo en la señal de audio del primer y segundo formador de haces. Con respecto a un entorno tridimensional, es preferible que la primera línea recta virtual y la segunda línea recta virtual se crucen y definan un plano que puede estar orientado de manera arbitraria.

[0024] De esta manera se proporcionan medios para captar el sonido de manera espacialmente selectiva, es decir, captar el sonido que se origina en una ubicación objetivo específica como si se hubiera instalado un micrófono de "punto" cercano en esta ubicación. En lugar de instalar, en realidad, este micrófono de punto, se puede simular su señal de salida mediante el uso de dos formadores de haces ubicados en posiciones distantes diferentes.

[0025] Estos dos formadores de haces no están ubicados cerca uno de otro, sino que están situados de tal manera que cada uno de ellos ejecute una adquisición de sonido direccional independiente. Los "haces" se superponen en un punto pretendido y sus salidas individuales se combinan posteriormente para formar una señal de salida final. A diferencia de otras estrategias posibles, la combinación de dos salidas individuales no requiere información ni conocimiento alguno sobre la posición de los dos formadores de haces en un sistema coordinado común. De este modo, la configuración general para la adquisición de micrófonos de punto virtuales comprende dos formadores de haces que operan de modo independiente, más un procesador de señales que combina ambas señales de salida individuales en la señal del "micrófono de punto" remoto.

[0026] En una realización, el aparato comprende un primer y un segundo formador de haces, por ejemplo, dos micrófonos espaciales y un generador de señales, por ejemplo, una unidad combinatoria, por ejemplo, un procesador para lograr la "intersección acústica". Cada micrófono espacial tiene una clara selectividad direccional, es decir que atenúa el sonido que se origina en sitios fuera de su haz, en comparación con el sonido que se origina en un punto

dentro de su haz. Los micrófonos espaciales funcionan independientemente uno de otro. La ubicación de los dos micrófonos espaciales, también flexibles por naturaleza, es elegida de tal manera que la ubicación espacial esté situada en la intersección geométrica de los dos haces. En una realización preferida, los dos micrófonos espaciales forman un ángulo de alrededor de 90 grados con respecto a la ubicación objetivo. La unidad combinada, por ejemplo, el procesador, puede no conocer la ubicación geométrica de los dos micrófonos espaciales o la ubicación de la fuente objetivo.

[0027] Según una realización, el primer formador de haces y el segundo formador de haces están dispuestos, con respecto a la ubicación objetivo, de tal manera que la primera línea recta virtual y la segunda línea recta virtual se crucen entre sí y de tal manera que se crucen en la ubicación objetivo con un ángulo de intersección de entre 30 grados y 150 grados. En una realización adicional, el ángulo de intersección es de entre 60 grados y 120 grados. En una realización preferida, el ángulo de intersección es de aproximadamente 90 grados.

[0028] En un ejemplo que no forma parte de la invención, el generador de señales comprende un filtro adaptativo que tiene una pluralidad de coeficientes de filtro. El filtro adaptativo está dispuesto para recibir la señal de audio del primer formador de haces. El filtro está adaptado para modificar la señal de audio del primer formador de haces dependiendo de los coeficientes de filtro para obtener una señal de audio filtrada del primer formador de haces. El generador de señales está configurado para ajustar los coeficientes de filtro del filtro dependiendo de la señal de audio del segundo formador de haces. El generador de señales puede estar configurado para ajustar los coeficientes de filtro de tal manera que se reduzca al mínimo la diferencia entre la señal de audio filtrada del primer formador de haces y la segunda señal de audio del segundo formador de haces.

[0029] En una realización, el generador de señales comprende una calculadora de intersecciones para generar la señal de salida de audio en el dominio espectral sobre la base de las señales de audio del primer y segundo formador de haces. Según una realización, el generador de señales puede comprender además un banco de filtro de análisis para transformar las señales del primer y segundo formadores de haces de un dominio del tiempo a un dominio espectral, y un banco de filtros de síntesis para transformar la señal de salida de audio de un dominio espectral a un dominio del tiempo. La calculadora de intersecciones puede estar configurada para calcular la señal de salida de audio en el dominio espectral sobre la base de la señal de audio del primer formador de haces que está representada en el dominio espectral y de la señal de audio del segundo formador de haces que está representada en el dominio espectral.

[0030] En una realización adicional, la calculadora de intersecciones está configurada para computar la señal de salida de audio en el dominio espectral sobre la base de una densidad espectral cruzada de las señales del primer y segundo formadores de haces y sobre la base de una densidad espectral de energía de la primera o la segunda señal de audio del formador de haces.

[0031] Según una realización, la calculadora de intersecciones está configurada para computar la señal de salida de audio en el dominio espectral empleando la fórmula

$$Y_1(k,n) = S_1(k,n) \cdot G_1(k,n), \text{ donde } G_1(k,n) = \sqrt{\frac{C_{12}(k,n)}{P_1(k,n)}}$$

en la que $Y_1(k, n)$ es la señal de salida de audio en el dominio espectral, en la que $S_1(k, n)$ es la señal de audio del primer formador de haces, en la que $C_{12}(k, n)$ es una densidad espectral cruzada de las señales del primer y segundo formadores de haces y en la que $P_1(k, n)$ es la densidad espectral de energía de la señal de audio del primer formador de haces, o
empleando la fórmula

$$Y_2(k,n) = S_2(k,n) \cdot G_2(k,n), \text{ donde } G_2(k,n) = \sqrt{\frac{C_{12}(k,n)}{P_2(k,n)}}$$

en la que $Y_2(k, n)$ es la señal de salida de audio en el dominio espectral, en la que $S_2(k, n)$ es la señal de audio del segundo formador de haces, en la que $C_{12}(k, n)$ es una densidad espectral cruzada de las señales del primer y segundo formadores de haces y en la que $P_2(k, n)$ es la densidad espectral de energía de la señal de audio del segundo formador de haces.

[0032] En otra realización, la calculadora de intersecciones está adaptada tanto para calcular la señal $Y_1(k, n)$ y $Y_2(k, n)$ y como para seleccionar la menor de ambas señales como señal de salida de audio.

[0033] En otra realización, la calculadora de intersecciones está configurada para computar la señal de salida

de audio en el dominio espectral empleando la fórmula

$$Y_3(k,n) = S_1 \cdot G_{34}(k,n), \text{ donde } G_{34}(k,n) = \sqrt{\frac{C_{12}(k,n)}{0.5\left(P_1(k,n) + P_2(k,n)\right)}}$$

en la que $Y_3(k, n)$ es la señal de salida de audio en el dominio espectral, en la que S_1 es la señal de audio del primer formador de haces, en la que $C_{12}(k, n)$ es una densidad espectral cruzada de la señal de audio del primer formador de haces, en la que $P_1(k, n)$ es la densidad espectral de energía de la señal de audio del primer formador de haces y en la que $P_2(k, n)$ es la densidad espectral de energía de la señal de audio del segundo formador de haces, o empleando la fórmula

$$Y_4(k,n) = S_2 \cdot G_{34}(k,n), \text{ donde } G_{34}(k,n) = \sqrt{\frac{C_{12}(k,n)}{0.5\left(P_1(k,n) + P_2(k,n)\right)}}$$

en la que $Y_4(k, n)$ es la señal de salida de audio en el dominio espectral, en la que S_2 es la señal de audio del segundo formador de haces, en la que $C_{12}(k, n)$ es una densidad espectral cruzada de las señales del primer y segundo formadores de haces, en la que $P_1(k, n)$ es la densidad espectral de energía de la señal de audio del primer formador de haces, y en la que $P_2(k, n)$ es la densidad espectral de energía de la señal de audio del segundo formador de haces.

[0034] En otra realización, la calculadora de intersecciones puede estar adaptada tanto para calcular la señal $Y_3(k, n)$ y $Y_4(k, n)$ como para seleccionar la menor de ambas señales como señal de salida de audio.

[0035] Según otro ejemplo que no forma parte de la invención, el generador de señales puede estar adaptado para generar la señal de salida de audio mediante la combinación de las señales del primer y segundo formadores de haces para obtener una señal combinada y ponderando la señal combinada con un factor de ganancia. La señal combinada puede ser ponderada, por ejemplo, en un dominio del tiempo, en un dominio de subbanda o en un dominio de Transformada Rápida de Fourier.

[0036] En un ejemplo adicional que no forma parte de la invención, el generador de señales está adaptado para generar la señal de salida de audio mediante la generación de una señal combinada de tal manera que el valor de densidad espectral de energía de la señal combinada sea igual al mínimo del valor de densidad espectral de energía de las señales del primer y segundo formadores de haces por cada mosaico de tiempo-frecuencia.

[0037] Realizaciones preferidas de la presente invención se explicarán con respecto a las figuras adjuntas en las cuales:

La figura 1 ilustra un aparato para capturar información de audio desde una ubicación objetivo según una realización,

La figura 2 ilustra un aparato según una realización que utiliza dos formadores de haces y una etapa de cómputo de la señal de salida,

La figura 3a ilustra un formador de haces y un haz del formador de haces que se dirige hacia una ubicación objetivo,

La figura 3b ilustra un formador de haces y un haz del formador de haces que muestra más detalles,

La figura 4a ilustra una configuración geométrica de dos formadores de haces con respecto a una ubicación objetivo según una realización,

La figura 4b ilustra la configuración geométrica de los dos formadores de haces de la figura 4a y tres fuentes de sonido, y

La figura 4c ilustra la configuración geométrica de los dos formadores de haces de la figura 4b y tres fuentes de sonido representadas en una ilustración más detallada,

La figura 5 ilustra un generador de señales según una realización,

La figura 6 ilustra un generador de señales según otra realización, y

La figura 7 es un gráfico de flujo que ilustra la generación de una señal de salida de audio sobre la base de una densidad espectral cruzada y de una densidad espectral de energía según una realización.

[0038] La figura 1 ilustra un aparato para capturar información de audio desde una ubicación objetivo. El aparato comprende un primer formador de haces 110 que está dispuesto en un entorno de grabación y que tiene una primera característica de grabación. Además, el aparato comprende un segundo formador de haces 120 que está dispuesto en el entorno de grabación y que presenta una segunda característica de grabación. Además, el aparato comprende un generador de señales 130. El primer formador de haces 110 está configurado para grabar la señal de audio de un primer formador de haces s_1 al dirigirse el primer formador de haces 110 hacia la ubicación objetivo con respecto a la primera característica de grabación. El segundo formador de haces 120 está configurado para grabar una señal de audio del segundo formador de haces s_2 al dirigirse el segundo formador de haces 120 hacia la ubicación objetivo con respecto a la segunda característica de grabación. El primer formador de haces 110 y el segundo formador de haces 120 están dispuestos de tal manera que una primera línea recta virtual, a la que se define pasando a través del primer formador de haces 110 y la ubicación objetivo, y una segunda línea recta virtual, que se define pasando por el segundo formador de haces 120 y la ubicación objetivo, no son paralelas entre sí. El generador de señales 130 está configurado para generar una señal de salida de audio s sobre la base de la señal de audio del primer formador de haces s_1 y de la señal de audio del segundo formador de haces s_2, de tal manera que la señal de salida de audio s refleje una cantidad relativamente mayor de información de audio originada en la ubicación objetivo en comparación con la información de audio que llega de la ubicación objetivo en las señales de audio del primer y segundo formador de haces s_1, s_2.

[0039] La figura 2 ilustra un aparato según una realización que utiliza dos formadores de haces y una etapa de cómputo de la señal de salida como parte común de las señales de salida individuales de los formadores de haces. Se representa un primer formador de haces 210 y un segundo formador de haces 220 para registrar las señales de audio del primer y segundo formadores de haces, respectivamente. Un generador de señales 230 realiza el cómputo de la parte común de la señal (una “intersección acústica”).

[0040] La figura 3a ilustra un formador de haces 310. El formador de haces 310 de la realización ilustrada en la figura 3a es un aparato para la adquisición direccionalmente selectiva de sonido espacial. Por ejemplo, el formador de haces 310 puede ser un micrófono direccional o una matriz de micrófono. En otra realización, el formador de haces puede comprender una pluralidad de micrófonos direccionales.

[0041] La figura 3a ilustra una línea curva 316 que encierra un haz 315. Todos los puntos de la línea curva 316 que define el haz 315 se caracterizan por el hecho de que un nivel de presión sonora predefinido que opera desde un punto de la línea curva da lugar al mismo nivel de señal de salida del micrófono correspondiente a todos los puntos de la línea curva.

[0042] Además, la figura 3a ilustra un eje principal 320 del formador de haces. El eje principal 320 del formador de haces 310 se define por el hecho de que un sonido con un nivel de presión sonora predefinido que se origina en un punto en cuestión del eje principal 320 da lugar a un primer nivel de señal de salida del formador de haces que es mayor o igual a un segundo nivel de señal de salida del formador de haces como resultado de un sonido con el nivel de presión sonora predefinido que se origina en cualquier otro punto que esté a la misma distancia del formador de haces que el punto en cuestión.

[0043] La figura 3b ilustra esto de forma más detallada. Los puntos 325, 326 y 327 están a una distancia igual d del formador de haces 310. Un sonido con un nivel predefinido de presión sonora que se origina en el punto 325 del eje principal 320 da lugar a un primer nivel de señal que sale del formador de haces que es superior o igual a un segundo nivel de señal de salida del formador de haces que es el resultado de un sonido con el nivel predefinido de presión sonora que se origina, por ejemplo, en el punto 326 o en el punto 327, que están a la misma distancia d del formador de haces 310 que el punto 325 del eje principal. En el caso tridimensional, esto significa que el eje principal indica el punto de una bola virtual con el formador de haces está situado en el centro de la bola, que genera el mayor nivel de señal de salida del formador de haces cuando un nivel predefinido de presión sonora se origina en el punto en comparación con cualquier otro punto de la bola virtual.

[0044] Volviendo a la figura 3a, también se ilustra una ubicación objetivo 330. La ubicación objetivo 330 puede ser una ubicación desde la cual se originan los sonidos que un usuario pretende grabar utilizando el formador de haces 310. Para ello, el formador de haces se puede dirigir a la ubicación objetivo para registrar el sonido pretendido. En este contexto, se considera que un formador de haces 310 se dirige a una ubicación objetivo 330, cuando el eje principal 320 del formador de haces 310 pasa a través de la ubicación objetivo 330. En ocasiones, la ubicación objetivo 330 puede ser una zona objetivo mientras que, en otros ejemplos, la ubicación objetivo puede ser un punto. Si la ubicación objetivo 330 es un punto, se considera que el eje principal 320 pasa a través de la ubicación objetivo 330 cuando el punto está situado en el eje principal 320. En la figura 3, el eje principal 320 del formador de haces 310 pasa a través de la ubicación objetivo 330 y, por lo tanto, el formador de haces 310 se dirige a la ubicación objetivo.

[0045] El formador de haces 310 tiene una característica de grabación que indica la capacidad del formador de haces para captar el sonido dependiendo de la dirección en la cual se origina el sonido. La característica de grabación del formador de haces 310 comprende la dirección del eje principal 320 en el espacio, la dirección, forma y las propiedades del haz 315, etc.

[0046] La figura 4a ilustra una configuración geométrica de dos formadores de haces, un primer formador de haces 410 y un segundo formador de haces 420, con respecto a una ubicación objetivo 430. Se ilustra un primer haz 415 del primer formador de haces 410 y un segundo haz 425 del segundo formador de haces 420. Además, la figura 4a representa un primer eje principal 418 del primer formador de haces 410 y un segundo eje principal 428 del segundo formador de haces 420. El primer formador de haces 410 está dispuesto de tal manera que se dirija hacia la ubicación objetivo 430, al pasar el primer eje principal 418 a través de la ubicación objetivo 430. Además, el segundo formador de haces 420 también se dirige a la ubicación objetivo 430, al pasar el segundo eje principal 428 a través de la ubicación objetivo 430.

[0047] El primer haz 415 del primer formador de haces 410 y el segundo haz 425 del segundo formador de haces 420 se cruzan en la ubicación objetivo 430, donde está ubicada una fuente objetivo que da salida a un sonido. Un ángulo de intersección del primer eje principal 418 del primer formador de haces 410 y el segundo eje principal 428 del segundo formador de haces 420 está indicado con α. Lo óptimo es que el ángulo de intersección α sea de 90 grados. En otras realizaciones, el ángulo de intersección es de entre 30 grados y 150 grados.

[0048] En un entorno tridimensional es preferible que el primer eje principal y el segundo eje principal virtual se crucen y definan un plano que puede estar orientado de manera arbitraria.

[0049] La figura 4b ilustra la configuración geométrica de los dos formadores de haces de la figura 4a, que también ilustra tres fuentes de sonido src1, src2, src3. Los haces 415, 425 de los formadores de haces 410 y 420 se cruzan en la ubicación objetivo, es decir la ubicación de la fuente objetivo src_3. Sin embargo, la fuente src_1 y la fuente src_2, están situadas únicamente en uno de los dos haces 415, 425. Se debe tener en cuenta que tanto el primero como el segundo formador de haces 410, 420 están adaptados para la adquisición de sonido direccionalmente selectiva y sus haces 415, 425 indican el sonido adquirido por ellos, respectivamente. Por consiguiente, el primer haz 425 del primer formador de haces indica una primera característica de grabación del primer formador de haces 410. El segundo haz 425 del segundo formador de haces 420 indica una segunda característica de grabación del segundo formador de haces 420.

[0050] En la realización de la figura 4b, las fuentes src_1 y src_2 representan fuentes perjudiciales que interfieren con la señal de la fuente pretendida src_3. Sin embargo, las fuentes src_1 y src_2 también pueden ser consideradas como componentes ambientes independientes captados por los dos formadores de haces. Idealmente, la salida de un aparato según una realización solo devolvería src_3 y al mismo tiempo suprimiría por completo las fuentes no deseadas src_1 y src_2.

[0051] Según la realización de la figura 4b, se emplean dos o incluso más dispositivos para la adquisición de sonido direccionalmente selectiva, por ejemplo, micrófonos direccionales, matrices de micrófonos y los correspondientes formadores de haces, para obtener una funcionalidad de “micrófono de puntos remotos”. Los formadores de haces adecuados pueden ser, por ejemplo, matrices o micrófonos altamente direccionales tales como micrófonos de cañón, y se pueden emplear las señales de salida de, por ejemplo, las matrices de micrófonos o los micrófonos altamente direccionales como señales de audio de los formadores de haces. Se utiliza la funcionalidad del “micrófono de puntos remotos” para captar solo el sonido que se origina en una zona limitada alrededor del punto.

[0052] La figura 4c ilustra esto de forma más detallada. Según una realización, el primer formador de haces 410 captura el sonido desde una primera dirección. El segundo formador de haces 420, que está situado a bastante distancia del primer formador de haces 410, captura el sonido desde una segunda dirección.

[0053] El primer y el segundo formador de haces 410, 420 están dispuestos de tal manera que se dirijan a la ubicación objetivo 430. En las realizaciones preferidas, los formadores de haces 410, 420, por ejemplo, dos matrices de micrófonos están distantes entre sí y miran al punto objetivo desde direcciones diferentes. Esto difiere del procesamiento tradicional por matrices de micrófonos, en que solo se utiliza una matriz única y sus diferentes sensores están ubicados en estrecha proximidad. El primer eje principal 418 del primer formador de haces 410 y el segundo eje principal 428 del segundo formador de haces 420 forman dos líneas rectas que no están dispuestas en paralelo, sino que, por el contrario, se cortan con un ángulo de intersección α. El segundo formador de haces 420 estaría colocado óptimamente con respecto al primer formador de haces, cuando el ángulo de intersección es de 90 grados. En las realizaciones, el ángulo de intersección es de al menos 60 grados.

[0054] El punto objetivo o la zona objetivo para la captación del sonido es la intersección de ambos haces 415, 425. La señal procedente de esta zona se deriva procesando las señales de salida de los dos formadores de haces 410, 420, de tal manera que se compute una “intersección acústica”. Esta intersección se puede considerar parte de la señal que es común/coherente entre las señales de salida de los dos formadores de haces individuales.

[0055] Tal concepto aprovecha tanto la direccionalidad individual de los formadores de haces como la coherencia entre las señales de salida de los formadores de haces. Esto difiere del procesamiento por matrices de micrófonos comunes, en que solo se utiliza una matriz única y sus diferentes sensores están dispuestos muy cerca unos de otros.

[0056] De esta manera, el sonido emitido es capturado/adquirido desde un sitio objetivo específico. Esto se diferencia de las estrategias que utilizan micrófonos distribuidos para estimar la posición de las fuentes de sonido, aunque no apuntan a una grabación realzada de las fuentes de sonido localizadas mediante la consideración de la salida de matrices de micrófonos distantes como se propone según las realizaciones.

[0057] Además de utilizar micrófonos altamente direccionales, los conceptos según las realizaciones pueden ser implementados tanto con formadores de haces clásicos como con filtros espaciales paramétricos. Si el formador de haces introduce distorsiones de amplitud y fase dependientes de la frecuencia, esto se sabría y sería tomado en cuenta para el cómputo de la "intersección acústica".

[0058] En una realización, un dispositivo, por ejemplo, un generador de señales, computa un componente de "intersección acústica". Un dispositivo ideal para computar la intersección daría la señal completa, en caso de estar presente una señal en las señales de ambos formadores de haces (por ejemplo, las señales de audio registradas por el primer y el segundo formadores de haces) y produciría una salida cero si una señal está presente solo en una o ninguna de las señales de audio de los dos formadores de haces. Se pueden obtener buenas características de supresión que garanticen también una eficiencia favorable del dispositivo, por ejemplo, determinando la ganancia de transferencia de una señal presente solo en la señal de audio de un formador de haces y ajustándola en relación con la ganancia de transferencia correspondiente a una señal presente en las señales de audio de ambos formadores de haces.

[0059] Las señales de audio de los dos formadores de haces s_1 y s_2 pueden ser consideradas como una superposición de una señal común objetivo filtrada, retardada y/o escalada s y señales de ruido/interferencia individuales, n_1 y n_2, de tal manera que

$$s_1 = f_1(s) + n_1$$

y

$$s_2 = f_2(s) + n_2$$

donde $f_1(x)$ y $f_2(x)$ son las funciones individuales de filtrado, retardo y/o escalado presentes para las dos señales. Por consiguiente, la tarea consiste en estimar s de $s_1 = f_1(s) + n_1$ y $s_2 = f_2(s) + n_2$. Para evitar ambigüedades, $f_2(x)$ se puede ajustar a identidad sin pérdidas de generalidad.

[0060] El "componente de intersección" puede ser implementado de diferentes maneras.

[0061] Según una realización, la parte común entre las dos señales se computa empleando filtros, por ejemplo filtros LMS (Media del Cuadrado Mínimo) adaptativos clásicos, ya que son comunes para la cancelación del eco acústico.

[0062] La figura 5 ilustra un generador de señales según una realización, en el cual se computa una señal común s a partir de las señales s_1 y s_2 mediante el uso de un filtro adaptativo 510. El generador de señales de la figura 5 recibe la señal de audio del primer formador de haces s_1 y la señal de audio del segundo formador de haces s_2 y genera la señal de salida de audio sobre la base de las señales del primer y segundo formadores de haces s_1 y s_2.

[0063] El generador de señales de la figura 5 comprende un filtro adaptativo 510. El filtro adaptativo 510 ejecuta un esquema clásico de procesamiento clásico de adaptación/optimización por error cuadrático medio, que se conoce gracias a la cancelación de eco acústica. El filtro adaptativo 510 recibe la señal de audio de un primer formador de haces s_1 y filtra la señal de audio del primer formador de haces s_1 para generar una señal de audio filtrada del primer formador de haces s como señal de salida de audio. (Otra notación adecuada para s sería $\tilde{s}$, aunque, para mejor legibilidad, a continuación, se hace referencia a la señal de salida en audio en el dominio del tiempo como "s"). El filtrado de la señal de audio del primer formador de haces s_1 se lleva a cabo sobre la base de coeficientes de filtro ajustables del filtro adaptativo 510.

[0064] El generador de señales de la figura 5 emite la señal de audio filtrada del primer formador de haces s. Además, la señal de salida de audio filtrada del formador de haces s es alimentada asimismo a una calculadora de diferencias 520. La calculadora de diferencias 520 también recibe la señal de audio del segundo formador de haces y calcula la diferencia entre la señal de audio filtrada del primer formador de haces s y la señal de audio del segundo formador de haces s_2.

[0065] El generador de señales está adaptado para ajustar los coeficientes de filtro del filtro adaptativo 510 de tal manera que se reduzca al mínimo la diferencia entre la versión filtrada de s_1 (=s) y s_2. Por consiguiente, se puede considerar que la señal s, es decir la versión filtrada de s_1, representa la señal de salida coherente buscada. En consecuencia, la señal s, es decir la versión filtrada de s_1 representa la señal de salida coherente pretendida.

[0066] En otra realización, se extrae la parte común entre las dos señales sobre la base de una métrica de coherencia entre las dos señales; véase, por ejemplo, la métrica de coherencia descrita en
[Fa03] C. Faller y F. Baumgarte, "Binaural Cue Coding – Parte II: Schemes y applications," IEEE Trans. on Speech y Audio Proc., vol. 11, no. 6, Nov. 2003.

[0067] Véase, además, la métrica de coherencia descrita en [Fa06] y [Her08].

[0068] Se puede extraer una parte coherente de las dos señales de las señales que están representadas en un dominio del tiempo, aunque también, y preferentemente, de las señales que están representadas en un dominio espectral, por ejemplo, un dominio de tiempo/frecuencia.

[0069] La figura 6 ilustra un generador de señales según una realización. El generador de señales comprende un banco de filtro de análisis 610. El banco de filtro de análisis 610 recibe la señal de audio de un primer formador de haces $s_1(t)$ y una señal de audio del segundo formador de haces $s_2(t)$. Las señales del primer y segundo formadores de haces $s_1(t)$, $s_2(t)$ están representadas en un dominio del tiempo; t especifica el número de muestras temporales de la señal del respectivo formador de haces. El banco de filtro de análisis 610 está adaptado para transformar las señales del primer y segundo formadores de haces $s_1(t)$, $s_2(t)$ de un domino del tiempo a un dominio espectral, por ejemplo, un dominio de tiempo-frecuencia, para obtener una primera $S_1(k, n)$ y una segunda $S_2(k, n)$ señal de audio en el dominio espectral del formador de haces. En $S_1(k, n)$ y $S_2(k, n)$, k especifica el índice de frecuencia y n especifica el índice de tiempo de la señal de audio respectiva del formador de haces. El banco de filtro de análisis puede ser cualquier tipo de banco de filtros de análisis como, por ejemplo, bancos de filtros de análisis de Transformada de Fourier de Corto Tiempo (STFT), bancos de filtros polifásicos, bancos de filtros Espejo en Cuadratura (QMF), aunque también bancos de filtros de análisis de Transformada de Fourier Discreta (DFT), Transformada de Coseno Discreta (DCT) y Transformada de coseno Discreta Modificada (MDCT). Mediante la obtención de una señal de audio en el dominio espectral del primer y segundo formadores de haces S_1 y S_2, se pueden analizar las características de las señales de audio del formador de haces S_1 y S_2 por cada marco de tiempo y por cada una de varias bandas de frecuencia.

[0070] Además, el generador de señales comprende una calculadora de intersecciones 620 para generar una señal de salida de audio en el dominio espectral.

[0071] Por añadidura, el generador de señales comprende un banco de filtros de síntesis 630 para transformar la señal de salida de audio generada desde un dominio espectral a un dominio del tiempo. El banco de filtros de síntesis 630 puede comprender, por ejemplo, bancos de filtros de síntesis de Transformada de Fourier de Corto Tiempo (STFT), bancos de filtros polifásicos, bancos de filtros de síntesis Espejo en Cuadratura (QMF), aunque también bancos de filtros de síntesis de Transformada de Fourier Discreta (DFT), Transformada de Coseno Discreta (DCT) y Transformada de coseno Discreta Modificada (MDCT).

[0072] A continuación, se explican las maneras posibles de computar la señal de salida de audio, por ejemplo, mediante la extracción de una coherencia. La calculadora de intersecciones 620 de la figura 6 puede estar adaptada para computar la señal de salida de audio en el dominio espectral según una o más de estas modalidades.

[0073] La coherencia extraída es una medida del contenido coherente común, a la vez que compensa por las operaciones de escalado y desplazamiento de fases. Véase, por ejemplo:

[Fa06] C. Faller, "Parametric Multichannel Audio Coding: Synthesis of Coherence Cues," IEEE Trans. on Speech y Audio Proc., vol. 14, n. º 1, Ene 2006;

[Her08] J. Herre, K. Kjörling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Rödén. W. Oomen, K. Linzmeier, K. S. Chong: "MPEG Surround – The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding", Journal of the AES, Vol. 56, N. º 11, noviembre de 2008, pág. 932–955

[0074] Una posibilidad para generar un cálculo estimativo de la parte de señal coherente de las señales del primer y segundo formadores de haces es aplicar los factores cruzados a una de las dos señales. Los factores cruzados pueden ser promediados en el tiempo. En este caso, se asume que el retardo relativo entre las señales del primer y segundo formadores de haces es limitado, por lo que es sustancialmente menor que el tamaño de ventana de los bancos de filtros.

[0075] A continuación, se explican, de forma detallada, las realizaciones de cálculo de la señal de salida de audio en el dominio espectral mediante la extracción de la parte de la señal común y empleando una estrategia basada en la correlación sobre la base de un cálculo explícito de una medida de la coherencia.

[0076] Las señales $S_1(k,n)$ y $S_2(k,n)$ denotan representaciones en el dominio espectral de las señales de audio del formador de haces donde k es un índice de frecuencia y n es un índice de tiempo. Por cada mosaico de tiempo-frecuencia específico (k,n) especificado por un índice de frecuencia específico k y un índice de tiempo específico n,

existe un coeficiente por cada una de las señales $S_1(k,n)$ y $S_2(k,n)$. De las dos señales de audio en el domino espectral de los formadores de haces $S_1(k,n)$, $S_2(k,n)$, se computa la energía del componente de intersección. Esta energía del componente de intersección se puede calcular, por ejemplo, determinando la magnitud de la densidad espectral de energía cruzada (CSD) $C_{12}(k,n)$ de $S_1(k,n)$ y $S_2(k,n)$:

$$C_{12}(k,n) = | E\{S_1(k,n) \cdot S^*_2(k,n)\} |$$

[0077] En este caso, el superíndice * denota el conjugado de un número complejo y E{} representa la expectativa matemática. En la práctica, se reemplaza el operador de expectativa, por ejemplo, por el suavizado temporal o de frecuencia del término $S_1(k,n) \cdot S^*_2(k,n)$, dependiendo de la resolución de tiempo/frecuencia del banco de filtros empleado.

[0078] La densidad espectral de energía (PSD) $P_1(k,n)$ de la señal de audio del primer formador de haces $S_1(k,n)$ y la densidad espectral de energía $P_2(k,n)$ de la señal de audio del segundo formador de haces $S_2(k,n)$ se pueden computar según las fórmulas:

$$P_1(k,n) = E\{|S_1(k,n)|^2\}$$
$$P_2(k,n) = E\{|S_2(k,n)|^2\}.$$

[0079] A continuación, se presentan realizaciones de las implementaciones prácticas del cómputo de la intersección acústica Y(k,n) a partir de las señales de audio de los dos formadores de haces.

[0080] Una primera manera de obtener una señal de salida se basa en la modificación de la señal de audio del primer formador de haces $S_1(k,n)$:

$$Y_1(k,n) = S_1(k,n) \cdot G_1(k,n), \text{ donde } G_1(k,n) = \sqrt{\frac{C_{12}(k,n)}{P_1(k,n)}} \quad (1)$$

[0081] Del mismo modo, se puede derivar una señal de salida alternativa de la señal de audio del segundo formador de haces $S_2(k,n)$:

$$Y_2(k,n) = S_2(k,n) \cdot G_2(k,n), \text{ donde } G_2(k,n) = \sqrt{\frac{C_{12}(k,n)}{P_2(k,n)}} \quad (2)$$

[0082] Para determinar la señal de salida, puede ser ventajoso limitar el valor máximo de las funciones $G_1(k,n)$ y $G_2(k,n)$ a un determinado valor umbral, por ejemplo, a uno.

[0083] La figura 7 es un gráfico de flujo que ilustra la generación de una señal de salida de audio sobre la base de una densidad espectral cruzada y una densidad espectral de energía según una realización.

[0084] En la etapa 710 se computa una densidad espectral cruzada $C_{12}(k, n)$ de las señales del primer y segundo formadores de haces. Por ejemplo, se puede aplicar la fórmula anteriormente descrita $C_{12}(k,n) = | E\{S_1(k,n) \cdot S^*_2(k,n)\} |$.

[0085] En la etapa 720, se computa la densidad espectral de energía $P_1(k, n)$ de la señal de audio del primer formador de haces. Por otro lado, también se puede utilizar la densidad espectral de energía de la señal de audio del segundo formador de haces.

[0086] Seguidamente, en la etapa 730, se computa una función de ganancia G1(k, n) sobre la base de la densidad espectral cruzada calculada en la etapa 710 y la densidad espectral de energía calculada en la etapa 720.

[0087] Por último, en la etapa 740, se modifica la señal de audio del primer formador de haces $S_1(k, n)$ para obtener la señal de salida de audio pretendida $Y_1(k, n)$. Si se ha calculado la densidad espectral de energía de la señal de audio del segundo formador de haces en la etapa 720, a continuación, se puede modificar la señal de audio del segundo formador de haces $S_2(k, n)$ para obtener la señal de salida de audio buscada.

[0088] Dado que ambas implementaciones tienen un único término de energía en el denominador, que se puede reducir dependiendo de la ubicación de la fuente sonora activa con respecto a los dos haces, es preferible emplear una ganancia que represente la relación entre la intensidad sonora correspondiente a la intersección acústica y la intensidad sonora general o media captada por los formadores de haces. Se puede obtener una señal de salida aplicando la fórmula

$$Y_3(k,n) = S_1 \cdot G_{34}(k,n), \text{ donde } G_{34}(k,n) = \sqrt{\frac{C_{12}(k,n)}{0.5(P_1(k,n) + P_2(k,n))}} \quad (3)$$

,

o aplicando la fórmula:

$$Y_4(k,n) = S_2 \cdot G_{34}(k,n), \text{ donde } G_{34}(k,n) = \sqrt{\frac{C_{12}(k,n)}{0.5(P_1(k,n) + P_2(k,n))}} \quad (4)$$

.

[0089] En ambos ejemplos descritos anteriormente, las funciones de ganancias asumen valores pequeños en caso de que el sonido registrado en las señales de audio del formador de haces no comprenda componentes de señal de la intersección acústica. Por otro lado, los valores de ganancia se acercan al obtenido si las señales de audio del formador de haces corresponden a la intersección acústica buscada.

[0090] Además, para asegurarse de que solo aparezcan en la señal de salida de audio componentes que correspondan a la intersección acústica (a pesar de la direccionalidad limitada de los formadores de haces utilizados) puede ser aconsejable computar la señal de salida final como señal más baja (por intensidad) de Y_1 e Y_2 (o Y_3 e Y_4), respectivamente. En una realización, se considera la señal Y_1 o Y_2 de las dos señales Y_1, Y_2 como señal más baja, que tiene la menor intensidad promedio. En otra realización, se considera la señal Y_3 o Y_4 como la señal más baja de ambas señales Y_3, Y_4, que tiene la menor intensidad promedio.

[0091] Además, existen otras maneras de calcular las señales de salida de audio que, a diferencia de lo descrito con respecto a las realizaciones anteriores, hacen uso tanto de las señales del primero como de las del segundo formador de haces S_1 y S_2 (en lugar de utilizar solo sus energías) combinándolas en una sola señal que a continuación es ponderada mediante el uso de las funciones de ganancia descritas. Por ejemplo, se pueden sumar las señales del primer y segundo formador de haces S_1 y S_2 y a continuación ponderar la señal sumatoria así obtenida utilizando una de las funciones de ganancia anteriormente descritas.

[0092] La señal de salida de audio en el dominio espectral S se puede convertir de nuevo de una representación en el tiempo/frecuencia a una señal temporal mediante el uso de un banco de filtros de síntesis (inverso).

[0093] En otro ejemplo que no forma parte de la invención, se extrae la parte común entre las dos señales procesando los espectros de magnitud de una señal combinada (por ejemplo, una señal sumada), por ejemplo, de tal manera que tenga la PSD (Densidad Espectral de Energía de la intersección (por ejemplo, mínima) de las señales de ambos formadores de haces (normalizadas). Las señales de entrada pueden ser analizadas en una forma selectiva del tiempo/frecuencia, como se describiera anteriormente, y se efectúa una presunción idealizada de que las dos señales de ruido son escasas y disociadas, es decir que no aparecen en el mismo mosaico de tiempo/frecuencia. En este caso, una solución sencilla consistiría en limitar el valor de densidad espectral de energía (PSD) de una de las señales al valor de la otra señal después de algún procedimiento adecuado de renormalización/alineamiento. Se puede suponer que el retardo relativo entre las dos señales es limitado para que sea sustancialmente menor que el tamaño de ventana del banco de filtros.

[0094] Aunque algunos aspectos se han descrito en el contexto de un aparato, es obvio que estos aspectos también representan una descripción del procedimiento correspondiente, en el cual un bloque o dispositivo corresponde a una etapa del procedimiento o a una característica de una etapa del procedimiento. De manera análoga, los aspectos descritos en el contexto de una etapa del procedimiento también representan una descripción de un bloque o elemento correspondiente o de una característica de un aparato correspondiente.

[0095] Una señal generada según las realizaciones anteriormente descritas puede ser almacenada en un medio de almacenamiento digital o se puede transmitir por un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión conectado por cable tal como Internet.

[0096] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden ser implementadas en hardware o en software. La implementación se puede realizar empleando un medio de almacenamiento digital, por ejemplo, un disco blando, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene almacenadas en la misma señales de control legibles electrónicamente, que cooperan (o tienen capacidad para cooperar) con un sistema informático programable de tal manera que se ejecute el procedimiento respectivo.

[0097] Algunos ejemplos que no forman parte de la invención comprenden un soporte de datos no transitorio que comprende señales de control legibles electrónicamente, con capacidad para cooperar con un sistema informático programable de tal manera que se ejecute uno de los procedimientos descritos en esta invención.

[0098] En general, las realizaciones de la presente invención pueden ser implementadas en forma de producto de programa informático con un código de programa, siendo el código de programa operativo para ejecutar uno de los procedimientos al ejecutarse el producto de programa informático en un ordenador. El código de programa puede ser almacenado, por ejemplo, en un soporte legible por una máquina.

[0099] Otras realizaciones comprenden el programa informático para ejecutar uno de los procedimientos descritos en esta invención, almacenado en un soporte legible por una máquina.

[0100] En otras palabras, una realización del procedimiento de la invención consiste, por lo tanto, en un programa informático que consta de un código de programa para realizar uno de los procedimientos descritos en esta invención al ejecutarse el programa informático en un ordenador.

[0101] Un ejemplo adicional que no forma parte de la invención es, por lo tanto, en un soporte de datos (o medio de almacenamiento digital, o medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para ejecutar uno de los procedimientos descritos en esta invención.

[0102] Un ejemplo adicional que no forma parte de la invención es, por lo tanto, un flujo de datos o una secuencia de señales que representa el programa informático para ejecutar uno de los procedimientos descritos en esta invención. El flujo de datos o la secuencia de señales pueden estar configurados, por ejemplo, para ser transferidos a través de una conexión de comunicación de datos, por ejemplo, a través de Internet.

[0103] Un ejemplo adicional que no forma parte de la invención comprende un medio de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurado o adaptado para ejecutar uno de los procedimientos descritos en esta invención.

[0104] Un ejemplo adicional que no forma parte de la invención comprende un ordenador en el que se ha instalado el programa informático para ejecutar uno de los procedimientos descritos en esta invención.

[0105] En algunas realizaciones, se puede utilizar un dispositivo lógico programable (por ejemplo, una matriz de puertas programables en el campo) para ejecutar algunas o todas las funcionalidades de los procedimientos descritos en esta invención. En algunas realizaciones, una matriz de puertas programables en el campo puede cooperar con un microprocesador para ejecutar uno de los procedimientos descritos en esta invención. Por lo general, los procedimientos son ejecutados preferentemente por cualquier aparato de hardware.

[0106] Las realizaciones anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y detalles descritos en esta invención han de ser evidentes para los expertos en la materia. La invención está limitada por el alcance de las reivindicaciones de patente inminentes.

Referencias

[BS01] J. Bitzer, K. U. Simmer: “Superdirective microphone arrays” in M. Brandstein, D. Ward (eds.): “Microphone Arrays – Signal Processing Techniques y Applications”, Capítulo 2, Springer Berlin, 2001, ISBN: 978–3–540–41953–2

[BW01] M. Brandstein, D. Ward: “Microphone Arrays – Signal Processing Techniques y Applications”, Springer Berlin, 2001, ISBN: 978–3–540–41953–2

[CBH06] J. Chen, J. Benesty, Y. Huang: “Time Delay Estimation in Room Acoustic Environments: un Overview”, EURASIP Journal on Applied Signal Processing, Article ID 26503, TOMO 2006 (2006)

[Pu106] Pulkki, V., ''Directional audio coding in spatial sound reproduction y stereo upmixing,'' en Acta de la 28a Conferencia Internacional de AES, págs. 251–258, Piteå, Suecia, 30 de junio – 2 de julio de 2006.

[DiFi2009] M. Kallinger, G. Del Galdo, F. Küch, D. Mahne, y R. Schultz–Amling, “Spatial Filtering using Directional Audio Coding Parameters,” en Proc. IEEE Int. Conf. on Acoustics, Speech, y Signal Processing (ICASSP), abril de 2009.

[Ea01] Eargle J. ''The Microphone Book'' Focal press 2001.

[Elk00] G. W. Elko: “Superdirectional microphone arrays” en S. G. Gay, J. Benesty (eds.): “Acoustic Signal Processing for Telecommunication”, Capítulo 10, Kluwer Academic Press, 2000, ISBN: 978–0792378143

[Fa03] C. Faller y F. Baumgarte, “Binaural Cue Coding – Parte II: Schemes y applications,” IEEE Trans. on Speech y Audio Proc., vol. 11, n. º 6, nov. 2003

[Fa06] C. Faller, “Parametric Multichannel Audio Coding: Synthesis of Coherence Cues,” IEEE Trans. on Speech y Audio Proc., vol. 14, n. º 1, enero de 2006

[Fa108] C. Faller: “Obtaining a Highly Directive Center Channel from Coincident Stereo Microphone Signals”, Acta 124a de la convención de AES, Ámsterdam, Países Bajos, 2008, Preimpresión 7380.

[Her08] J. Herre, K. Kjörling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Rödén. W. Oomen, K. Linzmeier, K. S. Chong: ''MPEG Surround – The ISO/MPEG Standard for Efficient y Compatible Multichannel Audio Coding'', Journal of the AES, Vol. 56, N. º 11, noviembre de 2008, págs. 932–955

[SBM01] K. U. Simmer, J. Bitzer, y C. Marro: “Post–Filtering Techniques” en M. Brandstein, D. Ward (eds.): “Microphone Arrays – Signal Processing Techniques y Applications”, Capítulo 3, Springer Berlin, 2001, ISBN: 978–3–540–41953–2

[Veen88] B. D. V. Veen y K. M. Buckley. “Beamforming: A versatile approach to spatial filtering”. IEEE ASSP Magazine, páginas 4–24, abril de 1988.

[Vi106] L. Villemoes, J. Herre, J. Breebaart, G. Hotho, S. Disch, H. Purnhagen, y K. Kjörling, “MPEG Surround: The Forthcoming ISO Standard for Spatial Audio Coding,” en 28a Conferencia Internacional de AES, Pitea, Suecia, junio de 2006.

REIVINDICACIONES

1. Un aparato para capturar sonido desde una ubicación objetivo ubicada en un entorno de grabación, que comprende:

un primer formador de haces (110; 210; 410) que está dispuesto en un entorno de grabación para la adquisición direccionalmente selectiva de sonido espacial y que tiene una directividad con un primer lóbulo **caracterizado por** un primer eje principal,
un segundo formador de haces (120; 220; 420) que está dispuesto en el entorno de grabación para la adquisición direccionalmente selectiva de sonido espacial y que tiene una directividad con un segundo lóbulo **caracterizado por** un segundo eje principal, y
un generador de señales (130; 230),
en el que el primer formador de haces (110; 210; 410) está configurado para generar una primera señal de audio de formador de haces y se coloca de tal manera que el primer lóbulo se dirija hacia la ubicación objetivo, y
en el que el segundo formador de haces (120; 220; 420) está configurado para generar una segunda señal de audio de formador de haces y se coloca de tal manera que el segundo lóbulo se dirija hacia la ubicación objetivo, y
en el que el primer formador de haces (110; 210; 410) y el segundo formador de haces (120; 220; 420) están dispuestos de tal manera que el primer eje principal y el segundo eje principal no son paralelos entre sí y se cruzan en la ubicación objetivo,
en el que el generador de señales (130; 230) está configurado para generar una señal de salida de audio sobre la base de la señal de audio del primer formador de haces y de la señal de audio del segundo formador de haces, en el que la señal de salida de audio comprende una parte común entre la primera y la segunda señales de audio del formador de haces,
caracterizado porque el generador de señales (130; 230) comprende una calculadora de intersecciones (620) para generar la señal de salida de audio en el dominio espectral en base a la primera y segunda señal de audio del formador de haces, y
en el que la calculadora de intersecciones (620) está configurada para calcular la señal de salida de audio en el dominio espectral mediante el cálculo de una densidad espectral cruzada de la primera y segunda señal de audio del formador de haces y mediante el cálculo de una densidad espectral de energía de la primera o la segunda señal de audio del formador de haces.

2. Un aparato según la reivindicación 1, en el que el primer eje principal y el segundo eje principal están dispuestos de tal manera que se crucen en la ubicación objetivo con un ángulo de intersección de tal manera que el ángulo de intersección esté entre 30 grados y 150 grados.

3. Un aparato según la reivindicación 2, en el que el primer eje principal y el segundo eje principal están dispuestos de tal manera que se crucen en la ubicación objetivo de forma que el ángulo de intersección sea de aproximadamente 90 grados.

4. Un aparato según una de las reivindicaciones 1 a 3, en el que el generador de señal (130; 230) comprende, además:

un banco de filtros de análisis (610) para transformar la primera y la segunda señales de audio del formador de haces de un dominio de tiempo a un dominio espectral, y
un banco de filtros de análisis (630) para transformar la señal de salida de audio de un dominio espectral a un dominio de tiempo,
en el que la calculadora de intersección (620) está configurada para calcular la señal de salida de audio en el dominio espectral basándose en la primera señal de audio del formador de haces que se representa en el dominio espectral y en la segunda señal de audio del formador de haces que se representa en el dominio espectral, en el que el cálculo se realiza por separado en varias bandas de frecuencia.

5. Un aparato según una de las reivindicaciones 1 a 4, en el que la calculadora de intersección (620) está configurada para calcular la señal de salida de audio en el dominio espectral empleando la fórmula

$$Y_1(k,n) = S_1(k,n) \cdot G_1(k,n), \quad \text{donde} \quad G_1(k,n) = \sqrt{\frac{C_{12}(k,n)}{P_1(k,n)}}$$

en la que Y_1 (k, n) es la señal de salida de audio en el dominio espectral, en la que S_1 (k, n) es la primera señal de audio del formador de haces, en la que C_{12} (k, n) es una densidad espectral cruzada de la primera y la segunda señal de audio del formador de haces, y en la que P_1 (k, n) es una densidad espectral de energía de la primera señal de audio del formador de haces, o empleando la fórmula

$$Y_2(k,n) = S_2(k,n) \cdot G_2(k,n), \quad donde \quad G_2(k,n) = \sqrt{\frac{C_{12}(k,n)}{P_2(k,n)}}$$

en la que Y_2 (k, n) es la señal de salida de audio en el dominio espectral, en la que S_2 (k, n) es la segunda señal de audio del formador de haces, en la que C_{12} (k, n) es una densidad espectral cruzada de la primera y la segunda señal de audio del formador de haces, y en la que P_2 (k, n) es una densidad espectral de energía de la segunda señal de audio del formador de haces.

6. Un aparato según una de las reivindicaciones 1 a 4, en el que la calculadora de intersección (620) está configurada para calcular la señal de salida de audio en el dominio espectral empleando la fórmula

$$Y_3(k,n) = S_1 \cdot G_{34}(k,n), \text{ donde } G_{34}(k,n) = \sqrt{\frac{C_{12}(k,n)}{0.5(P_1(k,n) + P_2(k,n))}}$$

en la que Y_3 (k, n) es la señal de salida de audio en el dominio espectral, en la que S_1 es la primera señal de audio de formador de haces, en la que C_{12} (k, n) es una densidad espectral cruzada de la primera señal de audio de formador de haces, en la que P_1 (k, n) es una densidad espectral de energía de la primera señal de audio del formador de haces, y en la que P_2 (k, n) es una densidad espectral de energía de la segunda señal de audio del formador de haces, o empleando la fórmula

$$Y_4(k,n) = S_2 \cdot G_{34}(k,n), \text{ donde } G_{34}(k,n) = \sqrt{\frac{C_{12}(k,n)}{0.5(P_1(k,n) + P_2(k,n))}}$$

en la que Y_4 (k, n) es la señal de salida de audio en el dominio espectral, en la que S_2 es la segunda señal de audio del formador de haces, en la que C_{12} (k, n) es una densidad espectral cruzada de la primera y la segunda señal de audio del formador de haces, en la que P_1 (k, n) es una densidad espectral de energía de la primera señal de audio del formador de haces, y en la que P_2 (k, n) es una densidad espectral de energía de la segunda señal de audio del formador de haces.

7. Un aparato según las reivindicaciones 5 o 6, en el que la calculadora de intersecciones (620) está adaptada para computar una primera señal intermedia según la fórmula

$$Y_1(k,n) = S_1(k,n) \cdot G_1(k,n), \quad donde \quad G_1(k,n) = \sqrt{\frac{C_{12}(k,n)}{P_1(k,n)}},$$

y una segunda señal intermedia según la fórmula

$$Y_2(k,n) = S_2(k,n) \cdot G_2(k,n), \quad donde \quad G_2(k,n) = \sqrt{\frac{C_{12}(k,n)}{P_2(k,n)}}$$

y en la que la calculadora de intersecciones (620) está adaptada para seleccionar la menor de la primera y segunda señales intermedias como la señal de salida de audio, o en la que la calculadora de intersecciones (620) está configurada para una tercera señal intermedia según la fórmula

$$Y_3(k,n) = S_1 \cdot G_{34}(k,n), \quad donde \quad G_{34}(k,n) = \sqrt{\frac{C_{12}(k,n)}{0.5(P_1(k,n) + P_2(k,n))}}$$

y una cuarta señal intermedia según la fórmula

$$Y_4(k,n) = S_2 \cdot G_{34}(k,n), \text{ donde } G_{34}(k,n) = \sqrt{\frac{C_{12}(k,n)}{0.5\left(P_1(k,n) + P_2(k,n)\right)}}$$

y en la que la calculadora de intersecciones (620) está adaptada para seleccionar la menor de la tercera y cuarta señales intermedias como la señal de salida de audio.

8. Un procedimiento para computar sonido de una ubicación objetivo en un entorno de grabación, que comprende:

generar una primera señal de audio de formador de haces mediante un primer formador de haces que está dispuesto en el entorno de grabación para la adquisición direccionalmente selectiva de sonido espacial y que tiene una directividad con un primer lóbulo **caracterizado por** un primer eje principal, en el que el primer formador de haces está posicionado de tal manera que el primer lóbulo esté dirigido hacia la ubicación objetivo,
generar una segunda señal de audio del formador de haces mediante un segundo formador de haces dispuesto en el entorno de grabación para la adquisición direccionalmente selectiva de sonido espacial y que tiene una directividad con un segundo lóbulo **caracterizado por** un segundo eje principal, en el que el segundo formador de haces se coloca de tal manera que el segundo lóbulo se dirige hacia la ubicación objetivo,
generar una señal de salida de audio basada en la primera señal de audio del formador de haces y en la segunda señal de audio del formador de haces, en el que la señal de salida de audio comprende una parte común entre la primera y la segunda señal de audio del formador de haces,
en el que el primer formador de haces (110; 210; 410) y el segundo formador de haces (120; 220; 420) están dispuestos de tal manera que el primer eje principal y el segundo eje principal no son paralelos entre sí y se cruzan en la ubicación de destino,
caracterizado porque la señal de salida de audio se genera en el dominio espectral mediante el cálculo de la primera y la segunda señal de audio del formador de haces, y
en el que la señal de salida de audio se calcula en el dominio espectral mediante el cálculo de una densidad espectral cruzada de la primera y la segunda señal de audio del formador de haces, y mediante el cálculo de una densidad espectral de energía de la primera o la segunda señal de audio del formador de haces.

9. Un producto de programa informático que comprende instrucciones que, cuando el programa es ejecutado por un ordenador, hace que el ordenador lleve a cabo el procedimiento de la reivindicación 8.

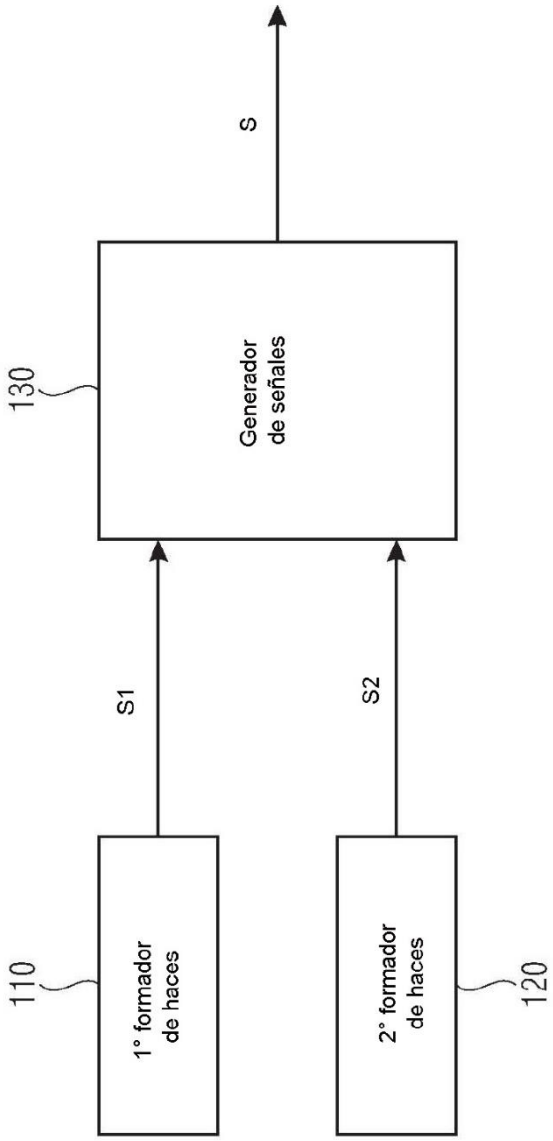
S
130
Generador de señales
S1
S2
110
1° formador de haces
2° formador de haces
120

FIG 1

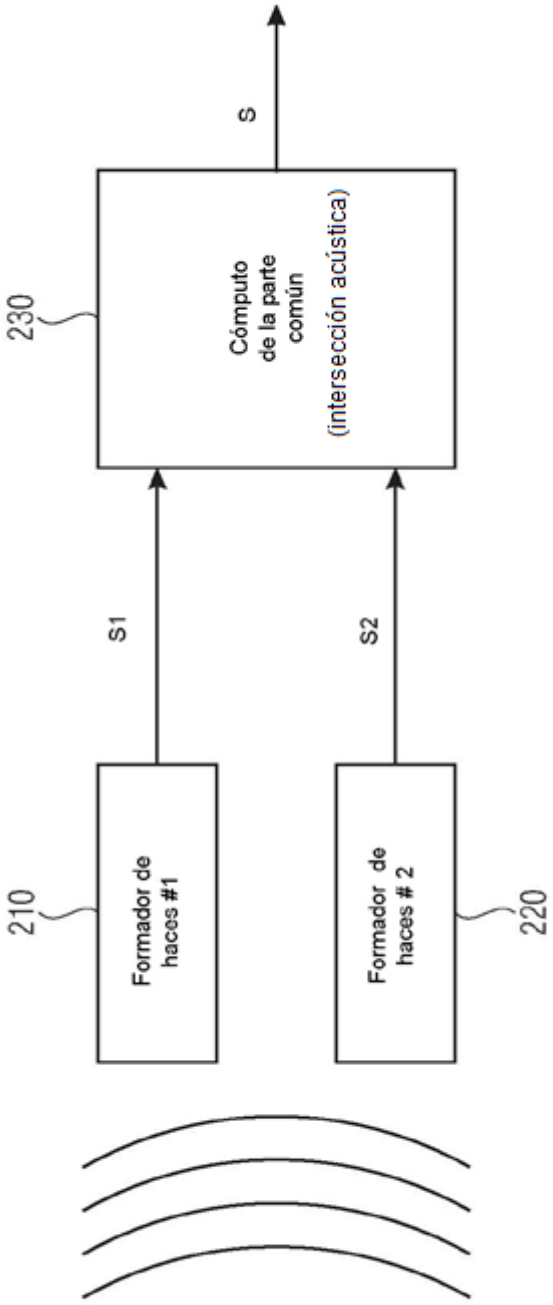
S
230
Cómputo
de la parte
común
(intersección acústica)
S1
S2
210
Formador de
haces #1
Formador de
haces # 2
220

FIG 2

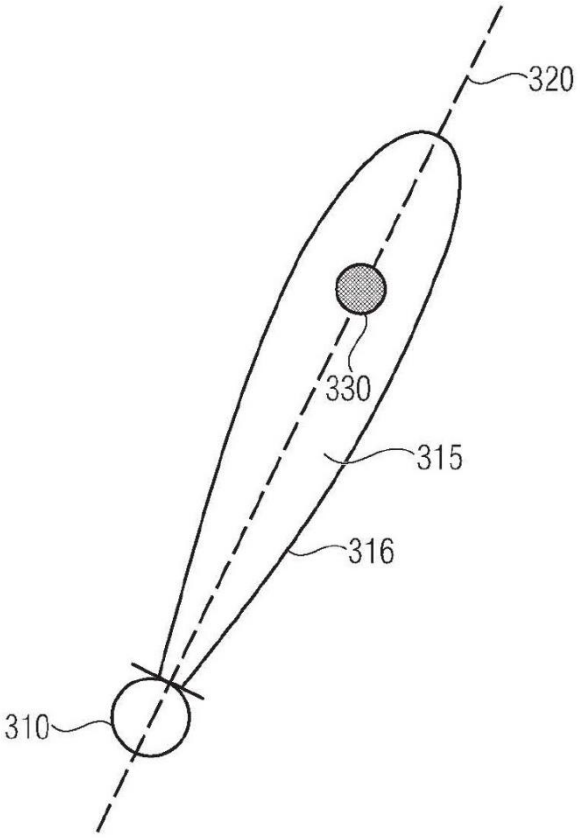

FIG 3A

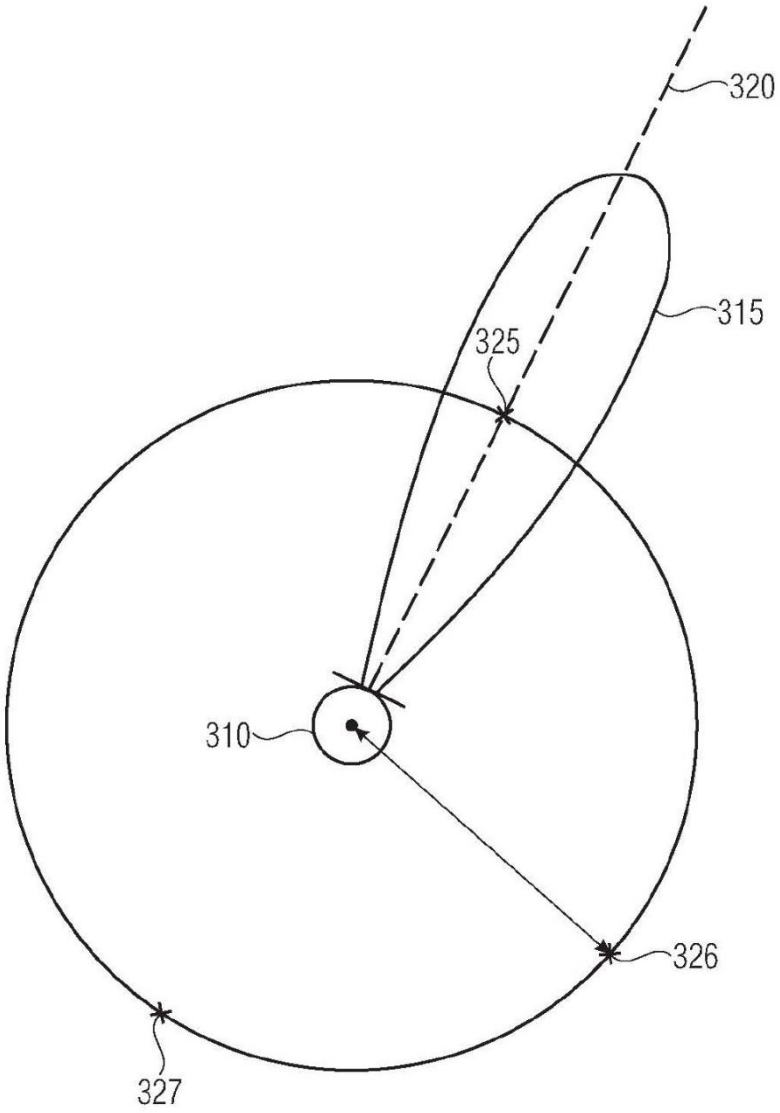

FIG 3B

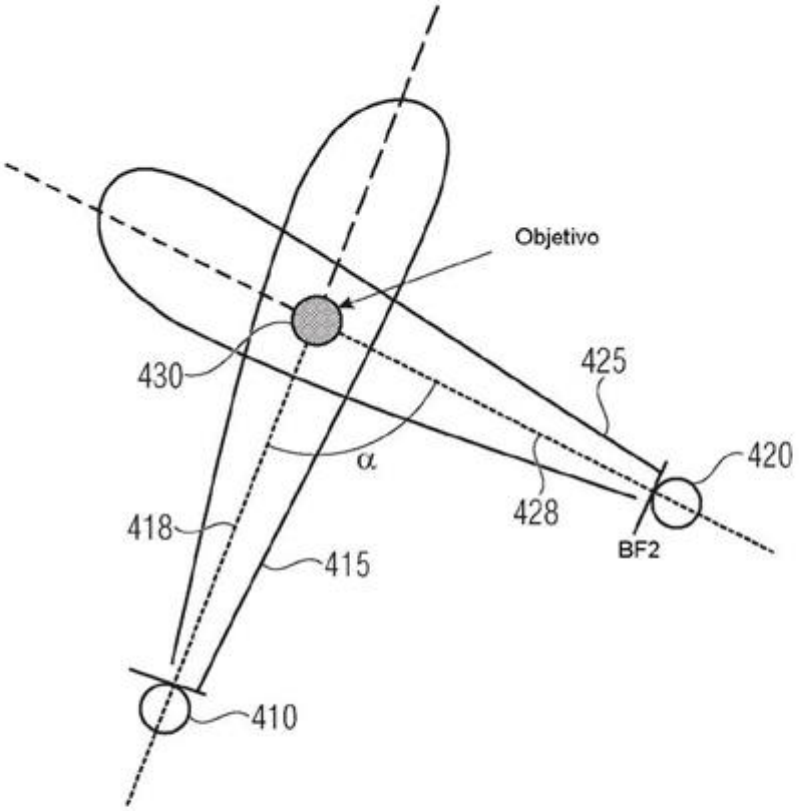

FIG 4A

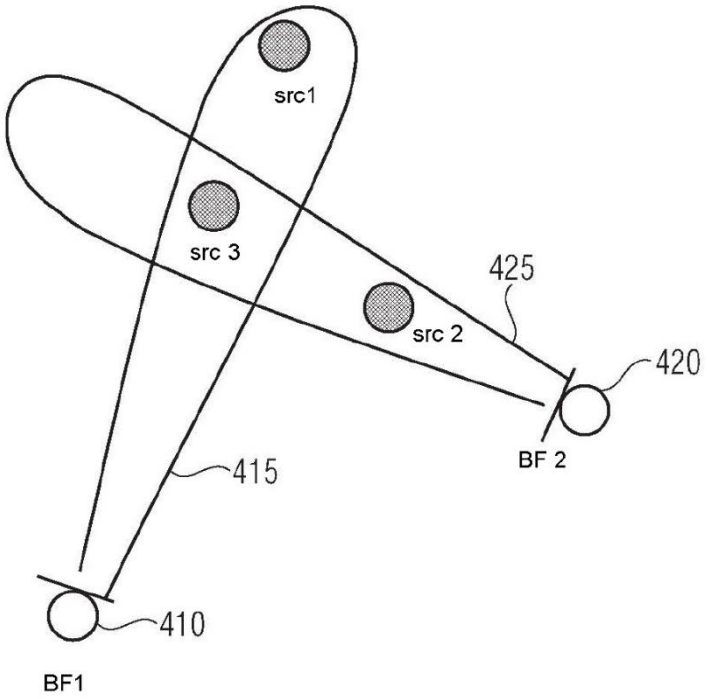

FIG 4B

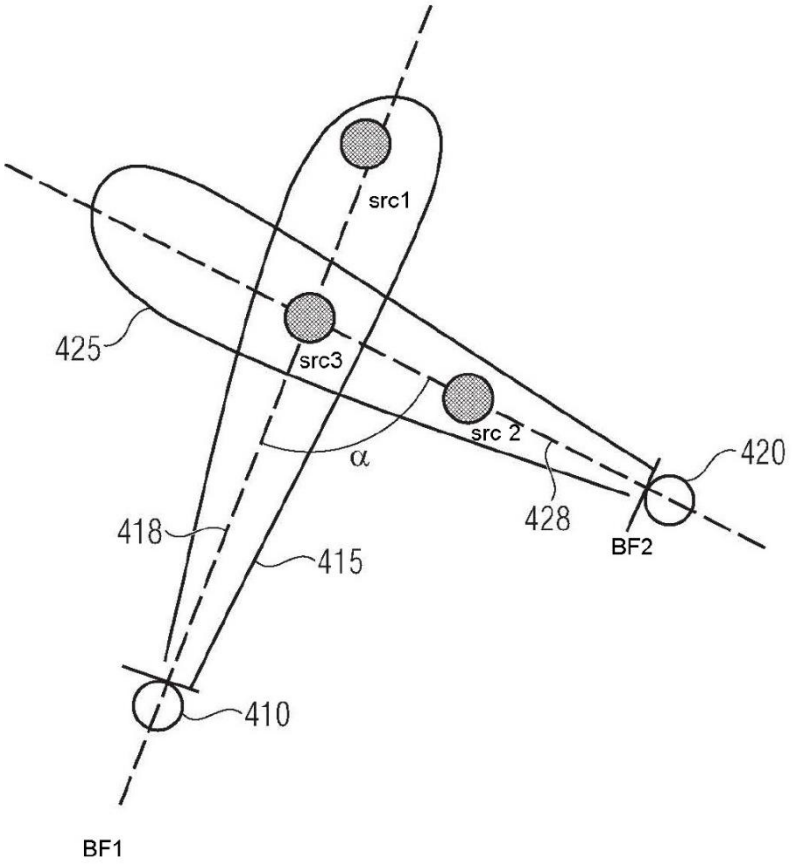

FIG 4C

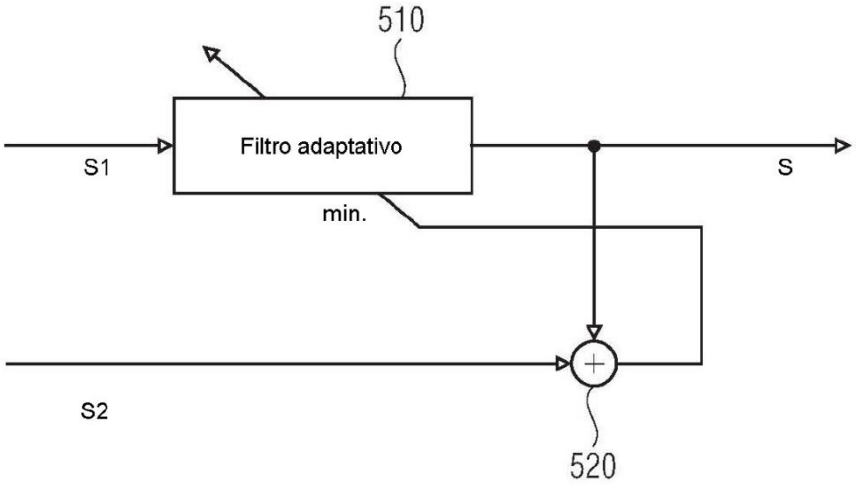

FIG 5

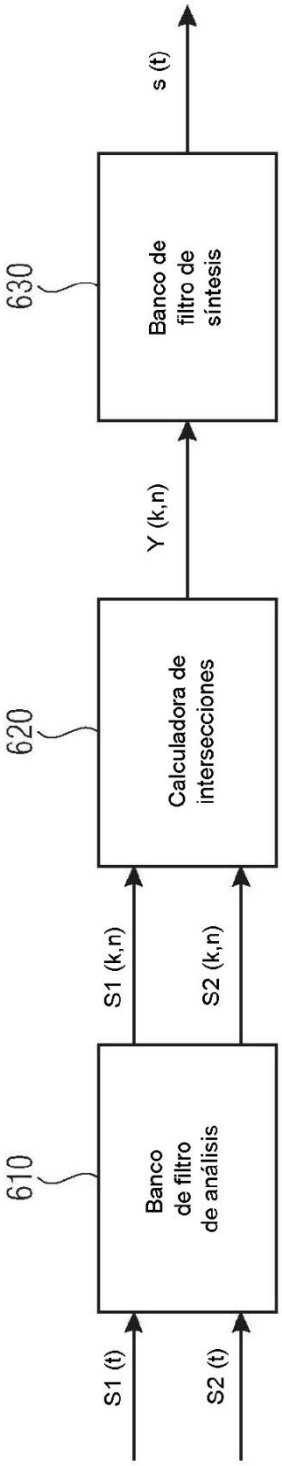
610
Banco de filtro de análisis
S1 (t)
S2 (t)
S1 (k,n)
S2 (k,n)
620
Calculadora de intersecciones
Y (k,n)
630
Banco de filtro de síntesis
s (t)

FIG 6

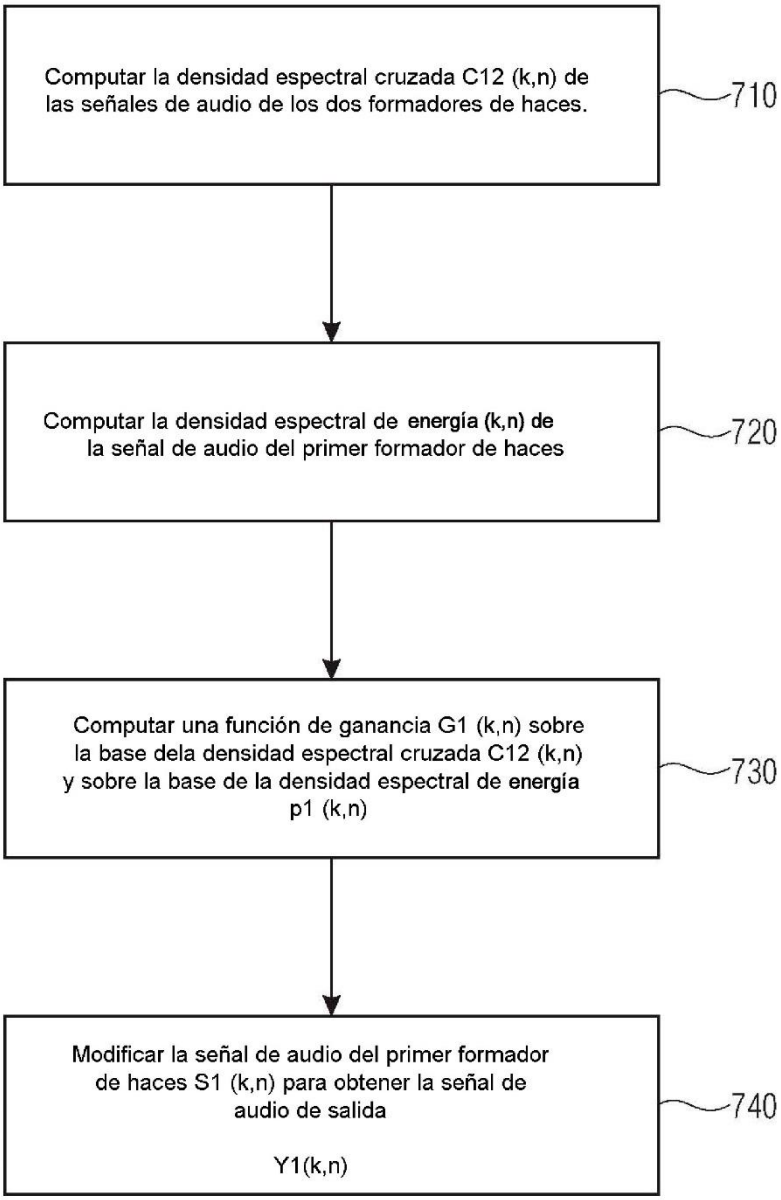
Computar la densidad espectral cruzada C12 (k,n) de las señales de audio de los dos formadores de haces.
710
Computar la densidad espectral de energía (k,n) de la señal de audio del primer formador de haces
720
Computar una función de ganancia G1 (k,n) sobre la base dela densidad espectral cruzada C12 (k,n) y sobre la base de la densidad espectral de energía p1 (k,n)
730
Modificar la señal de audio del primer formador de haces S1 (k,n) para obtener la señal de audio de salida
Y1(k,n)
740

FIG 7